

What is claimed is:

- 1 1. A method for extracting opinions about a subject of interest from a text
2 document having a plurality of sentences, the subject associated with a plurality of
3 features, the method comprising the steps of:
4 extracting from the document feature terms related to the features most
5 relevant to the subject;
6 for each sentence referring to a feature term, determining whether the
7 sentence includes an opinion polarity about the feature term; and
8 for each sentence referring to the subject, determining whether the sentence
9 includes an opinion polarity about the subject.
- 1 2. The method as recited in claim 1 further comprising the step of
2 determining an opinion skeleton for each opinion to provide support details for said
3 opinion.
- 1 3. The method as recited in claim 3, wherein the opinion skeleton
2 includes a feature term and an opinion term referring to said feature term.
- 1 4. The method as recited in claim 3, wherein the opinion skeleton
2 includes the subject and an opinion term referring to said subject.

1 5. The method as recited in claim 3, wherein the opinion skeleton
2 includes a feature term, an opinion term, and a relationship between the feature
3 term and the opinion term.

1 6. The method as recited in claim 3, wherein the opinion skeleton
2 includes the subject, an opinion term, and a relationship between the subject and
3 the opinion term.

1 7. The method as recited in claim 1, wherein the step of extracting the
2 feature terms includes the steps of:
3 determining a plurality of candidate feature terms associated the subject;
4 calculating a relevance score for each candidate feature term; and
5 identifying the most relevant feature terms from the candidate feature terms
6 based on the relevance scores.

1 8. The method as recited in claim 7, wherein each candidate feature term
2 is a definite noun phrase located at the beginning of a sentence.

1 9. The method as recited in claim 6, wherein the step of calculating a
2 relevance score is based on a likelihood ratio test.

1 10. The method as recited in claim 1, wherein the step of determining
2 whether the sentence includes an opinion polarity includes the steps of:
3 identifying opinion terms in the sentence using an opinion dictionary, each
4 entry in the dictionary having an opinion term, a part-of-speech tag, and an
5 associated opinion polarity;
6 for each sentence having a feature term and an opinion term, parsing the
7 sentence with an English parser to identify grammatical components in the sentence
8 and relationships between said components; and
9 identifying an opinion polarity associated with said feature term using the
10 opinion dictionary.

1 11. The method as in claim 10, wherein:
2 the grammatical components include verb phrases, subject phrases, object
3 phrases, complements, and prepositional phrases;
4 each feature term may have a modifier describing the feature term; and
5 the step of identifying an opinion polarity based on the opinion dictionary
6 includes the steps of:
7 for each sentence having a verb phrase where the verb phrase has no
8 matching entry in the opinion dictionary, assigning an opinion polarity of the modifier
9 of the feature term to the feature term, the opinion polarity of the modifier being
10 defined in the opinion dictionary.

1 12. The method as recited in claim 10, wherein the opinion polarity
2 associated with the feature term is identified based on an opinion rule.

1 13. The method as recited in claim 12, wherein:
2 the grammatical components include verb phrases, subject phrases, object
3 phrases, complements, and prepositional phrases; and
4 the step of identifying an opinion polarity based on the opinion rule base
5 includes the steps of:
6 for a sentence having a verb phrase, finding a matching entry in the
7 rule base; and
8 if there is a matching entry, assigning an opinion polarity to the feature
9 term as defined by the matching entry.

1 14. The method as recited in claim 12, wherein the rule base comprises a
2 plurality of rules each having a relationship term, a target of the opinion, and a
3 polarity of the opinion.

1 15. The method as recited in claim 12, wherein the rule base comprises a
2 plurality of rules each having a relationship term, a source of the opinion, and a
3 target of the opinion.

1 16. The method as recited in claim 15, wherein the target of the opinion is
2 a component of the sentence to which the opinion is to be assigned.

1 17. The method as recited in claim 15, wherein the source of the opinion is
2 a component of the sentence of which opinion polarity is to be assigned to the
3 target.

1 18. The method as recited in claim 1, wherein each feature term has a
2 “part-of” relationship with the subject.

1 19. The method as recited in claim 1, wherein each feature term has an
2 “attribute-of” relationship with the subject.

1 20. The method as recited in claim 1, wherein each feature term has an
2 “attribute-of” relationship with the associated feature.

1 21. A system for extracting opinions about a subject of interest from a text
2 document having a plurality of sentences, the subject associated with a plurality of
3 features, the system comprising:

4 means for extracting from the document feature terms related to the features
5 most relevant to the subject;

6 for each sentence referring to a feature term, means for determining whether
7 the sentence includes an opinion polarity about the feature term; and

8 for each sentence referring to the subject, means for determining whether the
9 sentence includes an opinion polarity about the subject.

1 22. A computer-program product for use with a computer for extracting
2 opinions about a subject of interest from a text document having a plurality of
3 sentences, the subject associated with a plurality of features, the computer-program
4 product comprising:

5 a computer-readable medium;

6 means, provided on the computer-readable medium, for extracting from the
7 document feature terms related to the features most relevant to the subject;

8 means, provided on the computer-readable medium, for each sentence
9 referring to a feature term, for determining whether the sentence includes an opinion
10 polarity about the feature term; and

11 means, provided on the computer-readable medium, for each sentence
12 referring to the subject, for determining whether the sentence includes an opinion
13 polarity about the subject.